

Dissecting the paternal founders of Mundari (Austroasiatic) speakers associated with the language dispersal in South Asia

Prajval Pratap Singh ^{1,2}, Shani Vishwakarma ², Gazi Nurun Nahar Sultana ³, Arno Pilvar ², Monika Karmin ^{2,4}, Siiri Rootsi ², Richard Villems ², Mait Metspalu ², Doron M Behar ², Toomas Kivisild ^{2,5}, George van Driem ^{6,7} and Gyaneshwer Chaubey ^{1,2}

¹ Cytogenetics Laboratory, Department of Zoology, Banaras Hindu University, Varanasi 221005, India

² Estonian Biocentre, Institute of Genomics, University of Tartu, Tartu 51010, Estonia

³ Centre for Advanced Research in Sciences (CARS), Genetic Engineering and Biotechnology Research Laboratory, University of Dhaka, Dhaka, 1000, Bangladesh.

⁴ Statistics and Bioinformatics Group, School of Fundamental Sciences, Massey University, Palmerston North 4410, New Zealand

⁵ Department of Human Genetics, Katholieke Universiteit Leuven, Leuven, 3000, Belgium.

⁶ Institut für Sprachwissenschaft, Universität Bern, 3012 Bern, Switzerland

⁷ Sydney Social Sciences and Humanities Advanced Research Centre, University of Sydney, Australia

Running title: The origin of Indian Austroasiatic speakers

***Correspondence to:**

Gyaneshwer Chaubey, PhD

Cytogenetics Laboratory

Department of Zoology

Banaras Hindu University

Varanasi, India- 211005

Ph. +919560375165

E mail: Gyaneshwer.chaubey@bhu.ac.in

Keywords: Austroasiatic, South Asia, Haplogroup O2, Mundari, Tibeto-Burman

Supplementary materials

Data availability

We used Y chromosome sequence data published in the Karmin et al. ¹. The sample id of two Indian samples used in the phylogeny are, GS000016803-ASM (Gond population) and GS000016804-ASM (Ho population). These sequences can be downloaded from European Nucleotide Archive (ENA; <http://www.ebi.ac.uk/ena/>) under accession number PRJEB8108 or data repository of the Estonian Biocentre (www.ebc.ee/free_data/chrY). The sequence trace files of the various markers have been submitted to the figshare database (DOI: 10.6084/m9.figshare.12771023). The details of each markers have been shown in the Supplementary Table 1.

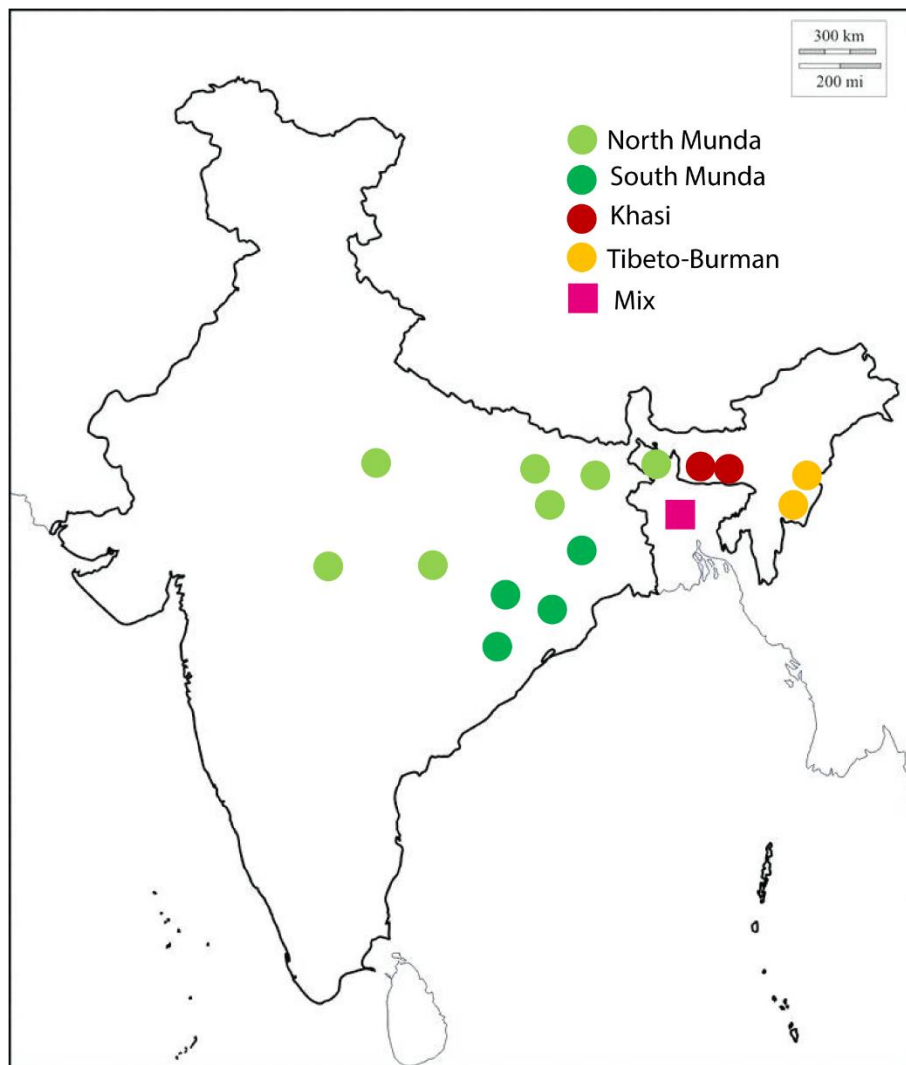
References

1. Karmin, M. *et al.* A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Res.* (2015).

Supplementary Figures and Tables

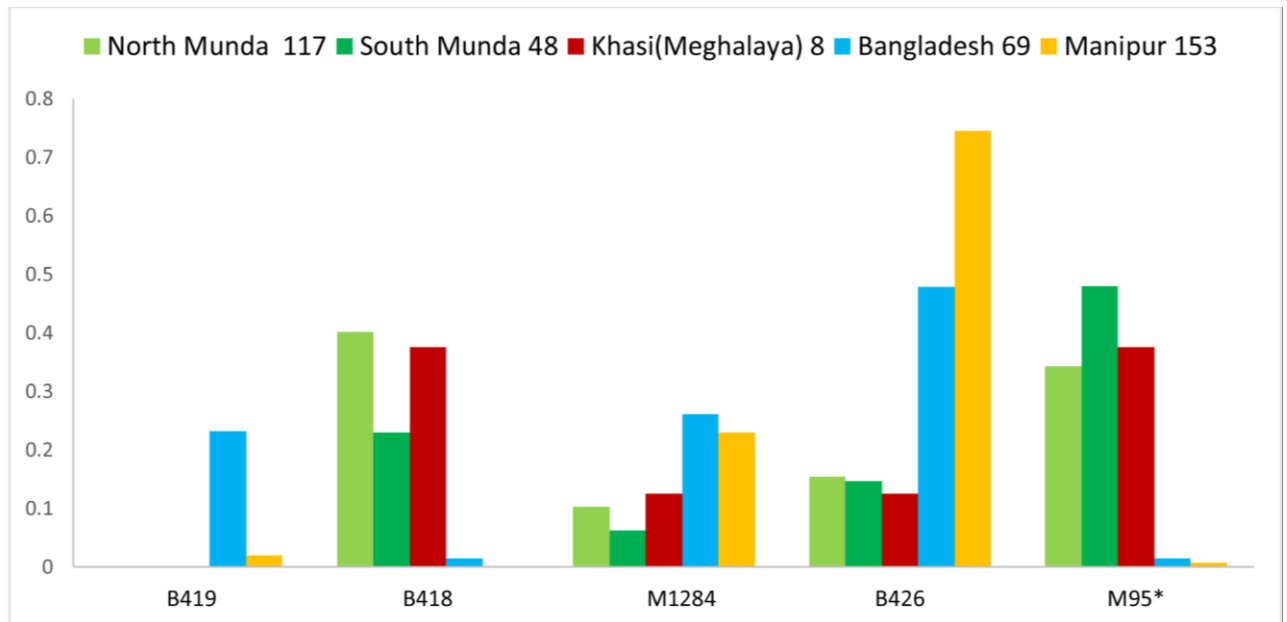
This file contains two figures and one table:

Supplementary Fig. 1. The sampling locations of various ethnic groups used in the present analysis. The detailed linguistic information of all the Bangladeshi samples genotyped were not present, therefore we have designated them as a 'mix' group.



Supplementary materials

Supplementary Fig. 2. The bar plot showing the frequency of newly defined and M95x(B418, B419, B426, M1284) branches amongst various ethnic groups of South Asia. This frequency is calculated out of the extant samples falling in haplogroup O2a-M95.



Supplementary materials

Supplementary Table 1: The details of the genetic markers used in this study.

Marker Name	Genomic Position (hg 18 chrY)	Forward primer	Reverse primer	PCR Product size	Enzymes	Cutting site	RFLP product
M1284 (eq. M1283)	g.6838496C>G	tttcacccaggctgcagt	agtacttaggccaggcacag	273bp	BseL1	5'..CCN ₅ ↓NN.. 3'	273bp 125+148 bp
B426	g.1509700A>G	gctcccaaacatttctgcta	taccagggttcaagatagc	215bp	-	-	-
B426eq	g.16794293T>C	gtgtccaggcatgtctttt	catccaactgcttccaact	383bp	Msp1	5'..CC↓GG..3'	383bp 220+163 bp
B424	g.18249536T>C	acgtgcaatctcacagttt	atgtcccagattgctgagaa	299bp	Taa1	5'..ACN↓GT..3'	299bp 121+178 bp
B418	g.22578932T>C	gggggtgagaaggtttgta	cagggagaaggaaactttca a	352bp	Bcl1	5'..T↓GAT..3' CA	352bp 286+66b p
B419	g.7750676A>C	gccagtgatatgacacaat	gcagatcacaatgtgaggag	262bp	Bsh123 61	5'..CG↓CG..3'	262bp 146+116 bp